

Claudia Freitas
Alexandre Rademaker (Eds.)

STIL 2015

**X Brazilian Symposium in Information and
Human Language Technology and Collocated
Events**

Proceedings of the Conference

**November 4 to 7, 2015.
Natal, Rio Grande do Norte.**

©2015 for the individual papers by the papers' authors. Copying permitted for private and academic purposes. Re-publication of material from this volume requires permission by the copyright owners.

Editors' addresses:

Pontifícia Universidade Católica do Rio de Janeiro
Rua Marquês de São Vicente, 225, Gávea
Rio de Janeiro, RJ - Brasil, 22451-900
claudiafreitas@puc-rio.br

IBM Research
Avenida Pasteur, 138 - Urca
Rio de Janeiro, RJ - Brazil, 22290-240
alexrad@br.ibm.com

X Brazilian Symposium in Information and Human Language Technology

This volume contains the papers presented at the X Symposium in Information and Human Language Technology (STIL 2015) and at the IV Workshop on Portuguese Description (JDP), held on November 4-7, 2015 in Natal, Brazil.

STIL is the bi-annual Language Technology event supported by the Brazilian Computer Society (SBC) and by the Brazilian Special Interest Group on Natural Language Processing (CE-PLN). The conference has a multidisciplinary nature and covers a broad spectrum of disciplines related to Human Language Technology, such as Linguistics, Computer Science, Psycholinguistics, Information Science, among others. It aims at bringing together both academic and industrial participants working on those areas.

The topics of interest centered around work in human language technology in general, such as Natural Language Resources & Tools, Corpus Linguistics, Text Classification, Sentiment Analysis and Opinion Mining, Information Extraction & Retrieval, Statistical and Machine Learning Methods, Natural language interfaces, Summarization, Terminology, Lexicology and Lexicography, to name a few.

We received 54 submissions from Brazil, Portugal, Norway, USA and Algeria. Each paper was reviewed by at least three members of the Program Committee, which had 62 members from 8 countries and 36 institutions. After a rigorous reviewing process, 15 papers were selected for oral presentation, and 10 papers were selected for poster presentation.

In addition, the conference featured invited talks by Adam Pease (IPsoft) and Lluís Padró (Universitat Politècnica de Catalunya), and a tutorial given by Lluís Padró.

We thank the authors for their submissions, the program committee for their hard work, invited speakers, SBC staff and the Local and General Chairs of STIL 2015.

November 2015

Cláudia Freitas
Alexandre Rademaker

Acknowledgments

The Program Committee chairs acknowledge the financial support to the conference provided by the Brazilian Computer Society (SBC), the Federal University of Rio Grande do Norte (UFRN) and the The North American Chapter of the Association for Computational Linguistics (NAACL). We thank the Program Committees of the X Brazilian Symposium in Information and Human Language Technology and Collocated Events for the reviews that they produced. Last but not least, we are grateful to the local organization led by Prof. Anne Magály de Paula Canuto (BRACIS 2015 General Chair) and Prof. Carlos Augusto Prolo (STIL Local Chair).

November 2015

Cláudia Freitas
Alexandre Rademaker

Program Chairs

Alexandre Rademaker (Fundação Getúlio Vargas & IBM Research, Brazil)
Maria Cláudia de Freitas (Pontifical Catholic University of Rio de Janeiro, Brazil)

Local Chair

Carlos A. Prolo (Federal University of Rio Grande do Norte, UFRN, Brazil)

Program Committee

Alberto Simões (University of Minho - UMinho, Portugal)
Aline Villavicencio (Federal University of Rio Grande do Sul - UFRGS, Brazil)
Andre Adami (University of Caxias do Sul - UCS, Brazil)
Antonio Branco (University of Lisbon - ULisboa, Portugal)
Arnaldo Candido Junior (Federal Technological University of Paraná - UTFPR, Brazil)
Carlos A. Prolo (Federal University of Rio Grande do Norte - UFRN, Brazil)
Carlos Ramisch (Aix Marseille Université - AMU, France)
Carmen Dayrell (University Nove de Julho - UNINOVE, Brazil)
Cassia Trojahn dos Santos (IRIT & University of Toulouse 2 - UTM2, France)
Cícero dos Santos (IBM Research, Brazil)
Christoph Treude (Federal University of Rio Grande do Norte - UFRN, Brazil)
Christopher Shulby (University of São Paulo - USP/ICMC, Brazil)
Clarissa Xavier (Pontifical Catholic University of Rio Grande do Sul - PUCRS, Brazil)
Daniel Lucrédio (Federal University of São Carlos - UFSCar, Brazil)
Daniel Muller (Federal University of Rio Grande do Sul - UFRGS, Brazil)
Diana Santos (Linguatca & University of Oslo - UiO, Norway)
Eraldo Fernandes (Federal University of Mato Grosso do Sul - UFMS, Brazil)
Erick Maziero (University of São Paulo - USP/ICMC, Brazil)
Ethel Schuster (Northern Essex Community College - NECC, USA)
Francis Bond (Nanyang Technological University - NTU, Singapore)
Geraldo Xexéo (Federal University of Rio De Janeiro - UFRJ, Brazil)
Gerard de Melo (Tsinghua University- THU, China)
Helena Caseli (Federal University of São Carlos - UFSCar, Brazil)
Heliana Mello (Federal University of Minas Gerais - UFMG, Brazil)
Heloisa Camargo (Federal University of São Carlos - UFSCar, Brazil)
Hugo Gonçalo Oliveira (University of Coimbra - UC, Portugal)
Ivandré Paraboni (University of São Paulo - USP/EACH, Brazil)
Jorge Baptista (University of Algarve - UAlg, Portugal)
José João Almeida (University of Minho - UMinho, Portugal)
Laura Alonso Alemany (National University Córdoba - UNC, Spain)
Leandro Mendonça de Oliveira (Brazilian Agricultural Research Corporation - EMBRAPA, Brazil)

Lucelene Lopes (Pontifical Catholic University of Rio Grande do Sul - PUCRS, Brazil)
Luciano Barbosa (IBM Research, Brazil)
Magali Duran (University of São Paulo - USP/ICMC, Brazil)
Marcelo Finger (University of São Paulo - USP/IME, Brazil)
Maria das Graças Nunes (University of São Paulo - USP/ICMC, Brazil)
Marilde Santos (Federal University of São Carlos - UFSCar, Brazil)
Márcio Dias (Federal University of Goiás - UFG, Brazil)
Milene Silveira (Pontifical Catholic University of Rio Grande do Sul - PUCRS, Brazil)
Norton Roman (University of São Paulo - USP/EACH, Brazil)
Pablo Mendes (IBM Research/Almaden, USA)
Palmira Marrafa (University of Lisboa - ULisboa, Portugal)
Paloma Moreda (University of Alicante - UA, Spain)
Paula Carvalho (European University, Portugal)
Paula Figueira Cardoso (University of São Paulo - USP/ICMC, Brazil)
Paulo Gomes (University of Coimbra - UC, Brazil)
Pedro Balage Filho (University of São Paulo - USP/ICMC, Brazil)
Renata Fortes (University of São Paulo - USP/ICMC, Brazil)
Renata Vieira (Pontifical Catholic University of Rio Grande do Sul - PUCRS, Brazil)
Roger Granada (Pontifical Catholic University of Rio Grande do Sul - PUCRS, Brazil)
Ronaldo Martins (UNDL Foundation - UNDLF, Brazil)
Ruy Milidiú (Pontifical Catholic University of Rio de Janeiro - PUC-Rio, Brazil)
Sandra Aluísio (University of São Paulo - USP/ICMC, Brazil)
Sara Candeias (Microsoft, Portugal)
Sérgio de Freitas (University of Brasília - UnB, Brazil)
Stella Tagnin (University of São Paulo - USP/FFLCH, Brazil)
Ted Pedersen (University of Minnesota - U of M, USA)
Thiago Pardo (University of São Paulo - USP/ICMC, Brazil)
Valéria Feltrim (State University of Maringá - UEM, Brazil)
Valeria de Paiva (Nuance Communications, USA)
Vera Lúcia Strube de Lima (Pontifical Catholic University of Rio Grande do Sul - PUCRS, Brazil)
Vlândia Pinheiro (University of Fortaleza - UNIFOR, Brazil)

Natural Language Processing Steering Committee

Vlândia C. M. Pinheiro, Universidade de Fortaleza - UNIFOR
Sandra M. Aluísio, Universidade de São Paulo - USP/São Carlos
Ariani Di Felippo, Universidade Federal de São Carlos - UFSCar
Cláudia Freitas, Pontifícia Universidade Católica do Rio de Janeiro - PUC-Rio
Valéria Delisandra Feltrim, Universidade Estadual de Maringá - UEM

Contents

I	Conference Papers	10
1	Invited Talks	11
	Numeric and Symbolic NLP: A Promising Engagement <i>Adam Pease</i>	12
	FreeLing: All you wanted to know and were afraid to ask <i>Lluís Padró</i>	13
2	Short Papers	14
	JCLext: A Java Tool for Compiling Finite-State Transducers from Full-Form Lexicons <i>Leonel F. de Alencar and Philipp B. Costa and Mardônio J. C. França and Alexander Ewart and Katuscia M. Andrade and Rossana M. C. Andrade</i>	15
	7x1-PT: um Corpus extraído do Twitter para Análise de Sentimentos em Língua Portuguesa <i>Sílvia M. W. Moraes and Isabel H. Manssour and Milene S. Silveira</i>	21
	Comparative Analysis between Notations to Classify Named Entities using Conditional Random Fields <i>Daniela Oliviera F. do Amaral and Maiki Buffet and Renata Vieira</i>	27
	Do Extrator de Conhecimento Coletivo à Ágora Virtual: desenvolvendo uma ferramenta para democracia participativa <i>Tiago Novaes Angelo and César José Bonjuani Pagan and Romis Ribeiro Faissol Attux and Ricardo Ribeiro Gudwin</i>	33
	Um novo corpo e os seus desafios <i>Diana Santos</i>	39
	Análise Automática de Coerência Textual em Resumos Científicos: Avaliando Quebras de Linearidade <i>Leandro Lago da Silva and Valéria Delisandra Feltrim</i>	45
	Anotação de corpus com a OpenWordNet-PT: um exercício de desambiguação <i>Cláudia Freitas and Livy Real and Alexandre Rademaker</i>	51
	Integrating support verb constructions into a parser <i>Amanda Rassi and Jorge Baptista and Nuno Mamede and Oto Vale</i>	57

CONTENTS

Extração de Alvos em Comentários de Notícias em Português baseada na Teoria da Centralização <i>Frank Willian Cardoso de Oliveira and Valéria Delisandra Feltrim</i>	63
Portal Min@s: Uma Ferramenta Geral de Apoio ao Processamento de Córpus de Propósito Geral <i>Arnaldo Candido Junior and Thiago Lima Vieira and Marcel Serikawa and Matheus Antônio Ribeiro Silva and Régis Zangirolami and Sandra Maria Alúcio</i>	69
PrepNet.Br: a Semantic Network for Prepositions. <i>Débora D. Garcia and Bento Carlos Dias da Silva</i>	75
3 Full Papers	80
Joint semantic discourse models for automatic multi-document summarization <i>Paula C. Figueira Cardoso and Thiago A. S. Pardo</i>	81
Building and Applying Profiles Through Term Extraction <i>Lucelene Lopes and Renata Vieira</i>	91
An Annotated Corpus for Sentiment Analysis in Political News <i>Gabriel Domingos de Arruda and Norton Trevisan Roman and Ana Maria Monteiro</i>	101
Campos Aleatórios Condicionais Aplicados à Detecção de Estrutura Retórica em Resumos de Textos Acadêmicos em Português <i>Alexandre C. Andreani and Valéria D. Feltrim</i>	111
Anotando um Corpus de Notícias para a Análise de Sentimentos: um Relato de Experiência <i>Mariza Miola Dosciatti and Lohann Paterno Coutinho Ferreira and Emerson Cabrera Paraiso</i>	121
Tesaurus Distribucionais para o Português: avaliação de metodologias <i>Rodrigo Wilkens and Leonardo Zilio and Eduardo Ferreira and Gabriel Gonçalves and Aline Villavicencio</i>	131
On Strategies of Human Multi-Document Summarization <i>Renata Tironi de Camargo and Ariani Di Felippo and Thiago A. S. Pardo</i>	141
Enriching entity grids and graphs with discourse relations: the impact in local coherence evaluation <i>Márcio de S. Dias and Thiago A. S. Pardo</i>	151
VerbLexPor: um recurso léxico com anotação de papéis semânticos para o português <i>Leonardo Zilio and Maria José Bocorny Finatto and Aline Villavicencio</i>	161
Novo dicionário de formas flexionadas do Unitex-PB: avaliação da flexão verbal <i>Oto A. Vale and Jorge Baptista</i>	171
Desambiguação de Homógrafos-Heterófonos por Aprendizado de Máquina em Português Brasileiro <i>Leonardo Hamada and Nelson Neto</i>	181
RePort - Um Sistema de Extração de Informações Aberta para Língua Portuguesa <i>Victor Pereira and Vlândia Pinheiro</i>	191

CONTENTS

Semi-Automatic Construction of a Textual Entailment Dataset: Selecting Candidates with Vector Space Models <i>Erick R. Fonseca and Sandra Maria Alúcio</i>	201
n-Gramas de Caractere como Técnica de Normalização Morfológica para Língua Portuguesa: Um Estudo em Categorização de Textos <i>Guilherme T. Guimarães and Marcus V. Meirose and Silvia M. W. Moraes</i>	211
II IV Jornada de Descrição do Português	221
4 Apresentação Oral	224
A utilização de atos de diálogo em sistemas de diálogo para dispositivos móveis. <i>Tiago Martins da Cunha and Daniel de França Brasil Soares</i>	225
Análise contrastiva da classificação sintático-semântica dos verbos locativos no Português do Brasil e no Português europeu <i>Roana Rodrigues and Jorge Baptista and Oto Vale</i>	233
A criação de um corpus de sentenças através de gramáticas livres de contexto <i>Tiago Martins da Cunha and Paulo Bruno Lopes da Silva</i>	241
Em direção à caracterização da complementaridade no corpus multidocumento CST-News <i>Jackson Souza and Ariani Di Felippo</i>	249
Explorando hierarquias conceituais para a seleção de conteúdo na sumarização automática multidocumento <i>Andressa C. I. Zacarias and Ariani Di Felippo</i>	257
Importância dos falsos homógrafos para a correção automática de erros ortográficos em Português <i>Magali Sanches Duran and Lucas Vinícius Avanço and Maria das Graças Volpe Nunes</i>	265
A inconsistência do tratamento dispensado às preposições pela gramática tradicional <i>Débora Domiciano Garcia and Bento Carlos Dias da Silva</i>	274
5 Apresentação por Pôster	283
As estratégias linguísticas e cognitivas que regem o Internetês - a escrita em rede - nos comentários do Facebook <i>Cristina Normandia and Maria Teresa Tedesco V. Abreu</i>	284
Estudos recentes sobre a detecção de contradição no processamento automático de línguas naturais <i>Denis Luiz Marcello Owa</i>	293
Transdutor de estados finitos para reconhecimento da nasalidade na pronúncia da variedade potiguar <i>Cid Ivan da Costa Carvalho</i>	301